

# ハイパフォーマンスチーム分析:

## 小サンプルデータのための探索的分析手法とその適用事例

### High-performance Team Analytics: A Method for Exploratory Analysis with Small Sample Data and Its Applications

仲間 大輔<sup>\*1</sup>  
Daisuke NAKAMA

宇野 渉<sup>\*1</sup>  
Wataru UNO

<sup>\*1</sup> 株式会社リクルートマネジメントソリューションズ HR Analytics & Technology Lab  
HR Analytics & Technology Lab, Recruit Management Solutions Co., Ltd.

Since enhancing team performance is an important issue in organizational management, there is a need to utilize human resource data to disentangle the characteristics of high-performance teams. In this study, we propose an analytic method to explore the relationship between all explanatory variables and team performance. The proposed method examines not only the direct effects of each variable, but also the moderation effects of all the combinations among variables, using both linear and nonlinear modeling. The effectiveness of the proposed method was verified with actual team data from three companies. In addition to interviewing the management of each company, we compared the results from the proposed method with those from the existing methods, and confirmed that the proposed method is superior in terms of both robustness and efficiency.

## 1. イントロダクション

「チーム」のパフォーマンスを高めることは組織マネジメントの重要な課題であり、人事データを活用してハイパフォーマンスなチームの特性を分析し、施策を検討するアプローチはニーズが高い。しかし、手法的な困難から実務的な活用は進んでいないのが現状である。そこで、本研究では、チーム改善のために重要となる要因を特定するための探索的な分析アプローチの開発を目的とする。

### 1.1 背景

組織内に存在する人事データの多くは属性情報やサーベイ結果等の個人単位データであり、チーム単位で分析するにはチームごとに基本統計量が集計されるが、チームを表す指標値の集計方法は多様に存在する(たとえば平均や標準偏差、最大値や最小値などはそれぞれにチームの各側面を表すとされる; Barrick 1998; Humphrey 2014)。そのため、検討対象となる変数の数(以下、「変数数」)が膨大になる傾向にある。また、一般に分析対象となるチーム数(以下、「n 数」)はそれほど大きくないこともあり、チーム分析データは、「横長」の形状となることが多い。このようなチームデータの特質から、a) 事前に立てた仮説についてデータで検証する、もしくは、b) 全変数について目的変数との関連を調べ(例えば相関係数を算出して)上位のものをピックアップする、といった分析手法が一般的となっている。

### 1.2 課題

これらの既存の手法には、しかし、それぞれに問題がある。既存手法 a(仮説検証型の分析)の場合、網羅性に問題があり、重要な変数を見逃してしまう可能性がある。また、既存手法 b(関連探索型の分析)の場合、小さな n 数であることから、結果のロバスト性に問題があること(e.g., 1 つデータを抜くと相関係数が大きく変化してしまう)、また、非線形効果や交互作用効果

といった、チームデータに特徴的に見られる性質(Mathieu 2008; Mathieu 2017)を考慮できていないという問題がある。

## 2. 提案手法

### 2.1 探索モデル

本研究では、上述の問題点の克服のため、解釈の容易性を重視して用いる変数数を限定しながら[豊田 1998]、網羅的な探索・抽出を行うための分析手法を提案する(Figure 1)。具体的には、全変数の直接的な関連だけでなく(直接要因モデル)、交互作用効果についても全変数の組み合わせについても(調整要因モデル)、線形・非線形モデリングの双方での分析を行う。


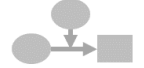
	直接要因モデル	調整要因モデル
		
	目的変数と説明変数が1対1	1個の目的変数と2個の説明変数
線形	パターン①: 線形相関	パターン③: 交互作用項を含む重回帰
非線形	パターン②: 決定木(深さ:2)	パターン④: 決定木(深さ:2)

Figure 1: 本研究の探索モデル

### 2.2 サンプリングと重要性の順位付け

また、n 数の少なさに由来する低ロバスト性の回避のために、全データから複数回サンプリングを行い、各サンプルデータで目的変数との関係性を表す「指標値」(e.g., 決定係数)を計算し、重要度を順位付けするという手法を取る(Figure 2)。

指標値としては、パターン①[直接要因×線形モデル]では相関係数を、パターン③[調整要因×線形モデル]では交互作用項の p 値を(効果量の代理指標として)、パターン②と④[非線形モデル]では決定係数を取った。

連絡先: 仲間大輔, リクルートマネジメントソリューションズ<sup>\*</sup> HAT Lab,  
070-3868-7001, daisuke\_nakama@recruit-ms.co.jp

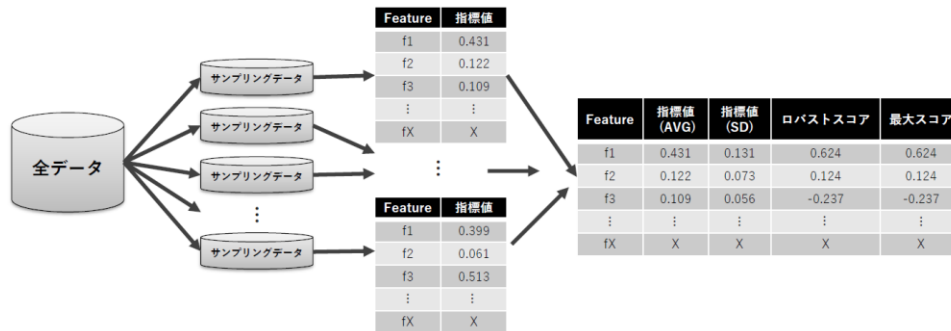


Figure 2: 本提案手法のスコア算出フロー

重要度の順位付けは、2通りの基準で行うことにする。まず、サンプリングデータ間で指標値のばらつきが小さく、安定して強い関係性が見られるという意味でロバストな関係を持つ変数の優先順位を高く設定することを目的とした基準（「ロバストスコア」基準）を考える。この基準は、サンプル毎に指標値が大きく上下する変数の優先度を下げようとする基準となる。サンプル毎に得られる指標値の平均に対して、ばらつき（標準偏差）の大きさが罰則項的な働きを持つように、各サンプルの指標値の 0 からの有意確率（ $p$  値）をスコア化して、それをを用いた順位付けを行った。

もう一つは、逆に、データの中に「隠れて」いるような、最大限の関係性の大きさを捉えようとする基準（最大スコア基準）である。これは、全データの中から、「例外」となるような観測値を除外した時に見られる、いわば潜在的な関係の大きさを見ようとしている。つまり、マジョリティのデータでは強い関係性があるにも関わらず、少数の例外的なチームのためにその関係性が埋没しているようなケースに対して高い優先順位を与えるような基準を考える。本研究では、各サンプルの指標値の最大値をスコア化し、順位付けした。

### 3. 検証事例

#### 3.1 検証の方法

提案手法の有効性について調べるため、実際の企業におけるチームデータを使用して、2通りの検証を行った。

まず、検証1として、既存手法 a（仮説検証型の分析）と提案手法との比較の観点から、分析結果の実務的な妥当性や、仮説としての事前の想定可能性を対象企業の担当者とのディスカッションを通じて検証した。さらに、検証2として、既存手法 b（全変数についての目的変数との関連の探索）との比較の観点から、提案手法と既存手法で抽出結果にどのような違いがあるのかを検証した。

#### 3.2 データ

検証には、 $n$  数の規模の異なる 3 社のデータを用いて分析を行った（Table 1）。

Table 1: 検証に使用した事例の概要

	A 社	B 社	C 社
n 数×変数数	20×1400	40×172	883×307
検討変数数 (直接要因モデル)	1400	172	307
検討組合せ数 (調整要因モデル)	979,300	111,156	46,971
業界・部門	製造・開発	製造・販売	製造・全社
目的変数	チーム生産性 (メンバー評定)	チーム業績 (売上達成率)	チームマネージャーへの評価
使用データ例	-組織調査 -多面調査 -適性検査	-チーム業績 -組織調査 -性格検査	-組織情報 -組織調査 -人事評価

なお、分析の際のパラメータとして、直接要因モデルではサンプル数 30、サンプルサイズ 90%、調整要因モデルではサンプル数 10、サンプルサイズ 90%とした。

#### 3.3 検証1

##### (1) 概要

提案手法により抽出された変数が、実際に実務場面において「重要」と言えるものであるのか、また、事前に仮説として想定し得るものであったのかを定性的に確認するため、各社の状況に対してブラインドの状態で行った分析を実施した上で、対象企業のマネジメント・人事担当者に対して分析結果のヒアリングを行った。

##### (2) 提案手法による分析の結果

データ形状や目的変数の種別が異なる 3 社のデータ全てに提案手法を適用することができ、それぞれに異なる結果が得られた（Table 2）。なお、重要度の順位付けにおいては、①～④の 4 つのパターンそれぞれにおいて、ロバストスコアに基づいて上位 10 変数、最大スコアに基づいて上位 5 変数を抽出した上で、実務的な妥当性の低いものを除外するという方法を取った。

Table 2: 提案手法で抽出された重要変数(群)の例

事例	項目例(抜粋)
A 社	職務遂行能力(自己知覚)の高さ(平均) 同僚の職務遂行能力の評価のばらつき(標準偏差)の少なさ
B 社	戦略理解度(平均) チームのリソース状況とマネージャーの性格特性の組み合わせ
C 社	マネジメント行動の多さ(平均)、偏りの小ささ(標準偏差) キャリア意識の高さ

##### (3) 検証の結果

担当者とのディスカッションの結果、3 社ともに、実務上感じている問題意識と整合的であり、提案された施策についても首肯できる結果とのフィードバックを得た。実感・妥当感のある要因抽出が行えていることを確認できたと同時に、要因間の優先順位付けを行う探索型の抽出方法の有効性を示唆する結果となった。

- 「組織を運営するにあたり、なんとなく課題はここかな？という感じはあったが、確認まではない状況だった。網羅的に分析して下さったため、抱いていた感覚に自信をもって、そしてデータをもって上の層に訴えることができる。」(A 社)
- 「たいへん参考になる。想定していた要因と必ずしも同じではないが、提案されている施策は納得感があり、今すぐにも取り組みたい。」(B 社)

- 「他の要因と比べた影響の強さは、結果を見て初めてわかった。感覚と近いところがあり、その裏付けになる。これから人事としてやっていかないとけない施策の後押しになる」(C社)

### 3.4 検証2

#### (1) 概要

提案手法と既存手法による変数の重要度の順位付けの差異について検証するため、もっとも単純なパターン①(直接要因×線形モデル)を例にした比較検証を行った。なお、「既存手法」としては、「全変数について目的変数との相関係数を計算し、その絶対値の大きさによって重要度を順位付けする」という方法を想定している。

#### (2) 抽出結果(順位付け)についての差異

既存手法と提案手法の間で、重要度の順位付けの違いをFigure 3にまとめた。

Figure 3から2つのことがわかる。第一に、順位乖離は、ロバストスコア・最大スコアのどちらの基準の場合でも、データ数が小さいA社とB社に顕著に見られ、データ数が大きいC社では比較的乖離は小さかった。これは、提案手法が小サンプルデータの場合に有効性が高いことを意味している。少データの場合には、サンプリング時に除外されるデータによって指標値(この場合では相関係数)が大きくばらつくため、そうした場合には、平均的な大きさとそのばらつきの両方を加味すること(ロバストスコア)や、潜在的な最大効果を考慮すること(最大スコア)の重要性が高まる。逆に言えば、十分なn数が確保できる場合には、提案手法の有用性はそれほど大きくないと言える。

第二に、手法間の順位乖離は、抽出対象に含める順位数(N)が小さい時に、より顕著になっていた。このことから、提案手法が探索の効率性を高めることがわかる。Nを十分に大きく取

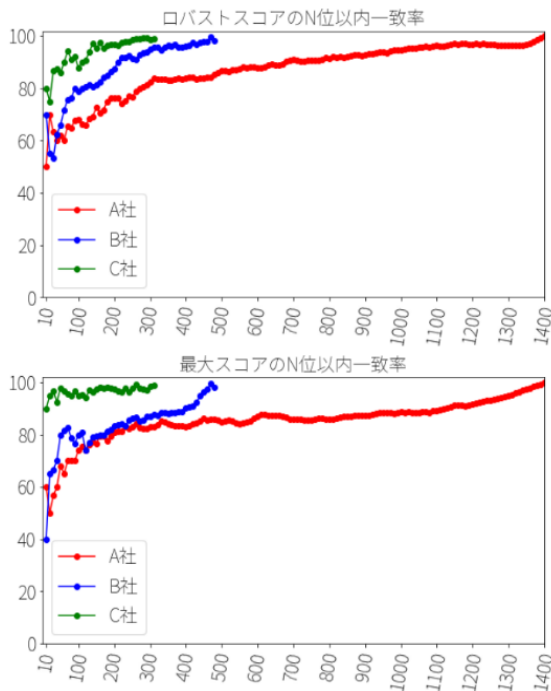


Figure 3: N位以内一致率

既存手法と提案手法のそれぞれで1~N位までの区間に含まれた変数の一致率。縦軸は一致率(%), 横軸はNを表す。上図が「ロバストスコア基準」による抽出順位での一致率, 下図が「最大スコア基準」による抽出順位での一致率。

れば、つまり、ほぼ全ての変数を考慮の対象とできる場合には、提案手法は既存手法と大きな違いはない。しかし、効率的な探索を行うために実務的な妥当性の検討の俎上に乗せる変数数を少なくする時に、提案手法は特に有効であった。

#### (3) 既存手法と提案手法の差異の傾向

次に、既存手法と提案手法との抽出内容の違いの詳細を確認するため、既存手法・提案手法の順位どちらかで50位以内に入った変数における、手法間の順位差(提案手法順位-既存手法順位)を調べた。提案手法により既存手法から順位が上昇したか下降したかの分類を見ると(Table 3)、ロバストスコア基準では上昇と下降が同程度に見られるのに対して、最大スコア基準では、順位の上昇の程度(順位差)が大きいことがわかった。つまり、ロバストスコア基準は、「既存手法では抽出できなかった変数の発見」(順位上昇)と同時に、「既存手法では抽出していた変数の除外」(順位下降)も同程度に行っているが、最大スコア基準は特に前者にフォーカスした選抜となっていた。

Table 3: 提案手法と既存手法による順位差

	ロバストスコア基準				最大スコア基準			
	順位下降		順位上昇		順位下降		順位上昇	
	個数	順位差(平均)	個数	順位差(平均)	個数	順位差(平均)	個数	順位差(平均)
A社	37	52.1	31	-52.9	32	28.4	32	-70.3
B社	34	24.1	33	-34.0	32	17.7	27	-32.2
C社	28	8.7	25	-14.0	20	3.2	17	-3.4

#### (4) 提案手法の優位性(1)——ロバストスコア基準

提案手法(ロバストスコア基準)で順位が上昇・下降した変数の詳細を見るために、B社を例にとって示した(Figure 4)。

ここでは、既存手法と提案手法による順位間ギャップについて特徴的なデータを3種類例示している。

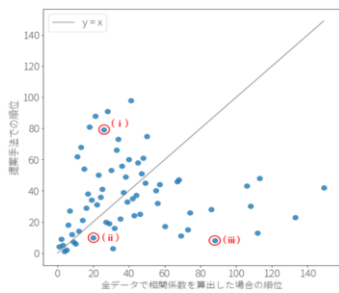
Figure 4(a)のiはy=xの傾向線から上側に離れた位置にあることから、通常の相関係数で算出した順位よりも提案手法により順位を下げたものである。逆に、iiはiと同程度の順位(どちらの相関係数も約0.34)であるが提案手法により順位が少しだけ上がったものである。各データの分布を見ると、iはiiに比べてサンプリングごとの相関係数の値が大きくばらつくことがわかる(Figure 4(b))。30回のサンプリングから得られた相関係数のばらつきはiは約0.08であるのに対し、iiは約0.044と小さい。

次に、iiiは、提案手法においてiiと同程度の順位となっている。iiiの相関係数は約-0.23と、iiの相関係数と絶対値で比較すると小さいにも関わらず、提案手法による順位はiiよりも高くなっている。これはiiiの分布に示すとおり、iiiはiiに比べてサンプリングごとの相関係数のばらつきが小さい分布(約0.03)のため、ロバストさを加味する提案手法において順位が上がったと言える。

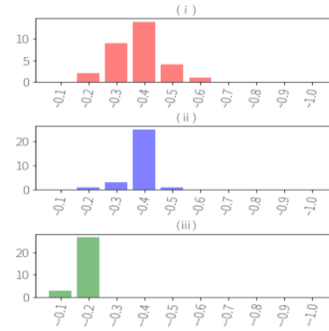
以上より、提案手法により、外れ値の影響で相関係数が高くなっているものは順位を下げ、逆に安定して高い相関係数が見られるものは順位を上げるため、ロバストに相関係数のある変数を、効率的にピックアップすることが可能になったと言える。

#### (5) 提案手法の優位性(2)——最大スコア基準

提案手法(最大スコア基準)による抽出の詳細を見るために、B社の事例における、「既存手法では抽出できなかった変数の発見」(順位上昇)に着目して図示したのがFigure 5である。

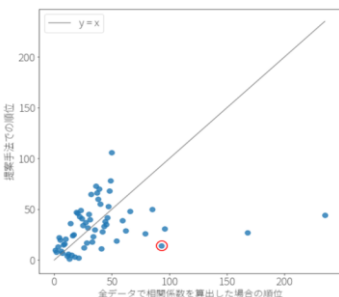


(a) 既存手法と提案手法の順位差  
縦軸は提案手法[ロバストスコア基準]での抽出順位, 横軸は既存手法での抽出順位.

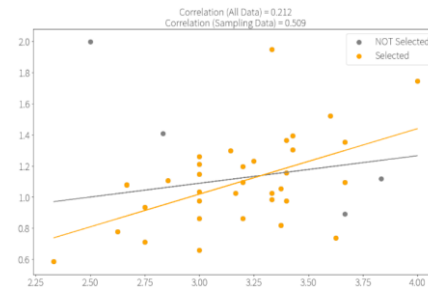


(b) 相関係数のヒストグラム  
Figure 4a上の3点 (i~iii) における, サンプルごと相関係数のヒストグラム

Figure 4: ロバストスコア基準により抽出された変数の妥当性



(a) 既存手法と提案手法の順位差  
縦軸は提案手法[最大スコア基準]での抽出順位, 横軸は既存手法での抽出順位.



(b) 着目データの散布図  
(a)上の着目点[赤丸]に関する, 目的変数(縦軸)と説明変数(横軸)の散布図.  
オレンジ点・線は相関係数が最大となったサンプリングデータとその回帰直線, 黒丸はその時に除外されたデータ, 黒線は全データの回帰直線.

Figure 5: 最大スコア基準により抽出された変数の妥当性

既存手法における順位と提案手法における順位との散布図である Figure 5 (a) において, 2 つの順位が同順位になる傾向線 ( $y=x$ ) よりも下側にある変数は, 提案手法により確認優先度が高くなったものであるが, 実際に赤丸で示したデータの分布を確認すると (Figure 5 (b)), 確かに一部の外れ値を除くことで高い相関係数となる変数であることが確認できた (0.212→0.509).

このように, 複数回のサンプリングごとに算出した複数の相関係数の最大値を見ることで, 大部分のデータでは相関が見られるが, 外れ値の影響で相関係数が小さくなってしまっている変数を提案手法によりピックアップすることが可能となったと言える.

#### 4. 考察と展望

本研究では, データ数は少ないが説明変数の数が膨大という特性を持つチームデータの探索的分析のための手法の開発を目的とし, 特定の形式のモデル (直接要因・調整要因モデル), 線形・非線形の双方の関係を網羅的に探索する手法を提案した (Figure 1).

探索の際には, サンプリングにより, 関連性のロバストさ (ロバストスコア基準) や潜在的な大きさ (最大スコア基準) による重要性評価を行った. さまざまに特徴の異なる3事例での検証の結果, 提案手法の有効性が確認できた. 網羅的な探索により発見的な意義があることに加え, 直感的な理解の容易さに由来する実務的な有用性があることもわかった (検証 1). さらに, 提案手法の有用性は,  $n$  数が多い場合よりも小さい場合に, また, 抽出する変数数を少なくしたい場合に, より有用であるという結論を得た (検証 2, Figure 3)

本提案手法の限界として, あくまで関連性の探索であり, 因果の検証とは異なる点が挙げられる (実際には, 追加分析で変数間のコントロール等も実施しているが, 本発表では割愛した). したがって, 本提案手法による分析結果は, マネジメントへのサジェスチョン目的で用いるべきであり, 因果推論手法の応用は今後に向けた課題となっている.

また, 分析に用いているデータ種別が事例によって異なるため, 本提案手法はチームプロセスに対する「一般解」を示すものではなく, 事例ごとに与えられた探索スペース内での探索となっていることにも留意が必要である. つまり, 複数の事例を横並びに比較することには意味がない. しかし, この点については, 事例の蓄積をすすめることで, メタ・アナリシスを可能にすることを目指していきたい.

さらに, 本提案手法は, 既存のデータを用いることを念頭においた分析手法である. 追加的なデータ取得を必要とせず, 既存データの分析を行えることは, 追加データ取得の実務上のハードルが高い実務場面に置いては大きなメリットとなるが, 事例によっては, 探索のためのデータが十分でない可能性がある. 今後は, 「分析目的に照らして最低限の必要データがあるか」という点を検討・確認するための枠組みの開発に取り組みたい.

#### 参考文献

- [Barrick 1998] Barrick, M R., Stewart, G.L., Neubert, M.J., and Mount, M.K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of applied psychology*, Vol. 83, No. 3, pp. 377-391, 1998.
- [Humphrey 2014] Humphrey, S.E., and Aime, F. Team micro-dynamics: Toward an organizing approach to teamwork. *Academy of management annals*, Vol. 8, pp. 443-503, 2014.
- [Mathieu 2008] Mathieu, J., Maynard, T., Rapp, T., and Gilson, L. Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of management*, Vol. 34, No. 3, pp. 410-476, 2008
- [Mathieu 2017] Mathieu, J, Hollenbeck, J.R., van Knippenberg, D., & Ilgen, D. R. A century of work teams in the *Journal of Applied Psychology*. *Journal of applied psychology*, Vol. 102, No. 3, pp. 452-467, 2017.
- [豊田 1998] 豊田秀樹. 共分散構造分析<入門編> —— 構造方程式モデリング——. 朝倉書店, 1998.